Different Discrete Wavelet Transforms Applied to Denoising Analytical Data

Chunsheng Cai and Peter de B. Harrington*,†

Department of Chemistry and Biochemistry, Center for Intelligent Chemical Instrumentation, Clippinger Laboratories, Ohio University, Athens, Ohio 45701-2979

Received May 25, 1998

Discrete wavelet transform (DWT) denoising contains three steps: forward transformation of the signal to the wavelet domain, reduction of the wavelet coefficients, and inverse transformation to the native domain. Three aspects that should be considered for DWT denoising include selecting the wavelet type, selecting the threshold, and applying the threshold to the wavelet coefficients. Although there exists an infinite variety of wavelet transformations, 22 orthonormal wavelet transforms that are typically used, which include Haar, 9 daublets, 5 coiflets, and 7 symmlets, were evaluated. Four threshold selection methods have been studied: universal, minimax, Stein's unbiased estimate of risk (SURE), and minimum description length (MDL) criteria. The application of the threshold to the wavelet coefficients includes global (hard, soft, garrote, and firm), level-dependent, data-dependent, translation invariant (TI), and wavelet package transform (WPT) thresholding methods. The different DWT-based denoising methods were evaluated by using synthetic data containing white Gaussian noise. The results of comparison have shown that most DWTs are very powerful methods for denoising and that the MDL and the TI methods are practical. The MDL criterion is the only method that can select a threshold for wavelet coefficients and select an optimal transform type. The TI method is insensitive to the wavelet filter so that for a variety of wavelet filters equivalent results were obtained. Savitzky–Golay and Fourier transform denoising results were used as reference methods. IR and HPLC data were used to compare denoising methods.

Wavelet analysis has become popular for signal processing in recent years, because it is an efficient method for data compression, fast computation, and noise reduction.[1] The Daubechies discrete wavelet transforms (DWTs) have been applied to data compression and noise reduction for multivariate calibration of near-infrared spectra.[2] The Daubechies DWT has been evaluated for smoothing electrospray mass spectra.[3] A method for fast PCA of data sets with high rank (i.e., greater than 10 000) using wavelet compression has been developed.[4] Two tutorials on wavelet transformation have been published.[5,6]

Experimental measurements usually contain noise that interferes with the interpretation of the data. High noise levels may be due to the instrumental instability, temperature fluctuation, etc., especially, when the measured signal is close to the detection limit. Denoising often is a preprocessing step before other analyses such as calibration or classification. Chemists use the DWT to denoise experimental data as an alternative method to the Fourier transform (FT) and Savitzky–Golay (SG). The commonly used wavelets are Haar, daublets, coiflets, and symmlets.Some wavelets are shown in Figure 1. WT decomposes the original domain data into a series of wavelets that have different scales and intensities. Mathematically, the computational procedures for these transforms are the same. The signal is multiplied by a transform matrix constructed from these filters. The results are permuted so that the detail and the smooth parts are separated, which is the first level transform. This procedure is repeated recursively on the smooth parts until the last level is reached at $\log_2 N$ steps. The wavelet computation is implemented by a sequence of special finite-length filtering steps. Daublets are derived from the Daubechies wavelet family. The daublets are designated by the filter lengths, which are integer values that typically range from 4 to 20 in steps of 2. The Haar transform is a special case of daublet 2. Coiflets usually consist of five filters, hence referred to as coiflet 1 to coiflet 5 with corresponding filter lengths that are multiples of six coefficients. Thus the coiflet 1 has six coefficients and coiflet 5 has 30 coefficients. The symmlet family has seven members that range from symmlet 4 to symmlet 10 with filter lengths that are multiples of two. A symmlet 4 has a filter length of 8.[7]

The DWT denoising procedure includes three steps. First, a data object with length of power of two is transformed into the wavelet domain. Second, some coefficients are selected and zero-filled or "shrunk" by some criterion. Third, the shrunk coefficients are inversely transformed to the original domain, which is the denoised data. The terms "shrinkage" and "shrunk" are used in the statistics literature. These terms refer to the attenuation of wavelet coefficient magnitude. The DWT based denoising methods can be classified as linear and nonlinear methods. The linear method truncates high frequency coefficients in wavelet domain. The assumption is that the signal is in the smooth part and the noise can be found in the detail part.This method may introduce large type I and type II errors. Type I error refers to the retention of noise components, and type II error refers to the loss of signal by the wavelet filter procedure. Therefore, this method alone is rarely used in practice as a denoising technique. Almost all denoising methods are nonlinear, which is to zero-fill or shrink those coefficients whose amplitudes are smaller than a threshold.

† E-mail: harring@helios.phy.ohiou.edu. Fax: 740-593-0148.

10.1021/ci980210j CCC: $15.00
Published on Web 09/25/1998

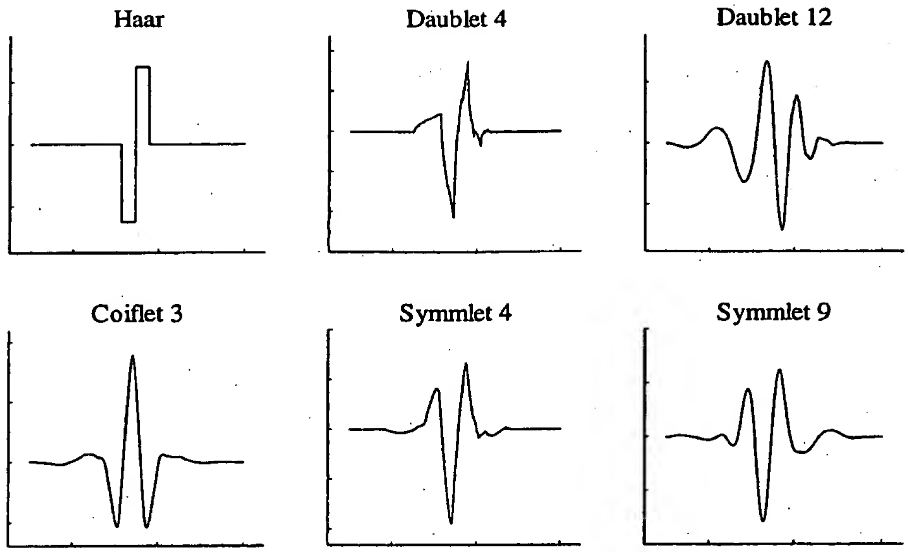

Figure 1. Some common wavelets.

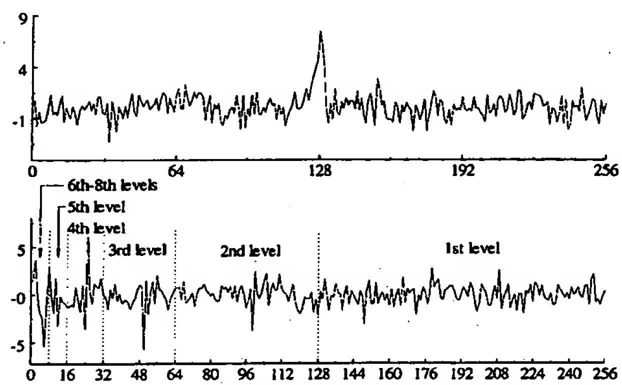

Figure 2. A Gaussian band at a *S/N* of 7 and the detail coefficients obtained from a symmlet 8 wavelet transform. The detail coefficients are obtained from several levels and are modified in the denoising procedure. Each level represents a recursive implementation of the WT.

The wavelet representation of a data object is a combination of many detail parts of the different transform levels, as given in Figure 2. The object on top is a Gaussian function at a signal-to-noise ratio (*S/N*) of 7. The figure below gives the detail wavelet coefficients obtained from the symmlet 8 transform. Each level recursively partitions the data into smooth and detail parts.

Detail coefficients below some threshold may be eliminated or shrunk in the denoising procedure. There are several approaches for defining a threshold criterion. A global threshold may be applied to all the wavelet coefficients.[8] A threshold may be defined for each level of the wavelet transform.[9] A data-dependent threshold criterion can also be used, which is a special case of the level-dependent threshold.[10,11] There are many other threshold criteria and methods that may be applied to wavelet denoising.

Problems arise when there are so many kinds of DWT denoising methods. In practice, selection of wavelet family and filter length is important. The selection may be guided by empirical rules applied to data size and signal continuity. The typical way is to visually inspect the data first, and if the data are kind of discontinuous, Haar or other sharp wavelet functions are applied, otherwise a smoother wavelet such as daublet 12 is employed. However, as shown in the following discussion session, even two similar wavelets may give significantly different denoising results. Furthermore, the threshold and the thresholding method must be selected for denoising. In this study, methodologies for selecting DWT type and the thresholds among the DWT-based denoising methods are compared. Synthetic data are used to evaluate the denoising methods. The results from these evaluations serve as a guide for the denoising problem. The traditional denoising methods of SG[12−14] and FT[15,16] are used as reference methods to evaluate the efficacy of the DWT methods.

THEORY

Denoising of experimental data can be viewed as a problem of nonparametric regression, in which a signal is recovered from a noisy signal. The goal of denoising is to obtain an estimate of the signal and remove the noise components. Among the three steps of DWT denoising, the second step is the most important. This step consists of determining a threshold and the treatment of the wavelet coefficients that are below this threshold. For DWT denoising, three aspects should be considered: selecting a DWT type, selecting a threshold, and applying the threshold to the wavelet coefficients.

1. Selecting Wavelet Type. Theoretically, there exists an infinite set of wavelet transforms, but the Haar, daublets, coiflets, and symmlets are widely used for signal processing. Among the 22 wavelet types, selecting the best wavelet type for specific data is difficult. As the results of this study will demonstrate, even two similar wavelets may give very different denoising results. Implementing all of these transforms and visually choosing the best denoising result is inefficient and subjective with regards to the scientist's bias. Therefore, selecting a wavelet filter that is matched to the data is a key step for wavelet transform denoising. Among the different denoising methods, only MDL can select filter type, which will be discussed in detail in the following sessions.

2. Selecting Threshold. There are four common threshold selection methods: universal, minimax, Stein's unbiased estimate of risk (SURE), and minimum description length (MDL). The universal threshold is defined by

$$t = \sigma \times \sqrt{2 \times \ln(N)} \tag{1}$$

for which N is the length of data array, and σ is the standard deviation of the noise.[8,17] For most real data, σ is unknown, but can be estimated as s. The first detail part of the wavelet coefficients χ_i can be used to estimate the noise by

$$s \approx \frac{\text{median}(|x_i|)}{0.6745} \tag{2}$$

for which s is the noise estimate.[17]

The minimax criterion gives a table of the threshold values for given data sizes that is based on calculations of the minimax risk bound for the wavelet estimate. Minimax thresholds were first introduced for soft thresholding (see below for the thresholding methods).[18] These threshold values are smaller in magnitude than the universal threshold values. Recently, minimax thresholds for hard, firm, and non-negative garrote thresholding have been derived.[19,20] Minimax thresholds optimize the risks for the worst cases, and therefore they are relatively conservative. This method estimates the noise level in the data using eq 2 and is biased toward retaining signal at the cost of retaining noise.

SURE is used to obtain an unbiased estimate of the variance between the filtered and unfiltered data. SURE is defined as

$$\text{SURE}(t,x) = N - 2 \times M_{(i:|x_i| \le t)} + \sum_{i=1}^{N}(|x_i| \wedge t)^2 \tag{3}$$

for which t is the candidate threshold, x_i is the wavelet coefficient, N is the data size, and M is the number of the data points less than t.[8,21] The t that yields the minimum SURE value is selected as the threshold value. The last term in the SURE function determines the residual energy after thresholding ($|x_i| \wedge t$ is the minimum value between $|x_i|$ and t). This criterion was originally developed for level-dependent soft thresholding. The SURE criterion can be applied to other thresholding methods. A modification of SURE threshold for global thresholding, called SPINSURE, was proposed by combining the SURE and cycle-spinning technique (see below).[22]

The MDL criterion is defined by

$$\text{MDL}(k^*,m^*) = \min\left(\frac{3}{2}k \log(N) + \frac{N}{2}\log(\sum(x_m^2 - x_{mk}^2))\right) \tag{4}$$

for which k is the number of largest coefficients that are retained, m designates the filter type, x_m for wavelet coefficients from transform type m, x_{mk} for the k largest coefficients in amplitude, and k^* and m^* are the optimized values.[23] The corresponding wavelet coefficient at k^* is assigned as the threshold. The $3/2k \log (N)$ term is a penalty function, which is proportional to the number of retained wavelet coefficients. The second log term characterizes the residual energy, which is the error between the reconstructed signal and the original noisy signal. Note this unique method not only picks a threshold but also a filter type. Neither the SURE nor the MDL criteria require an estimate of the noise level s.

3. Thresholding Methods. Thresholding methods refer to the ways of applying a threshold to the wavelet coefficients, i.e., how to modify the wavelet coefficients. Traditional thresholding methods all transformed coefficients whose magnitudes are below the threshold. There are other means to modify the coefficients. Because DWTs are multilevel transforms and the transformed coefficients come from different levels as shown in Figure 2, different thresholds may be applied to each different level. In DWT-based denoising family, cycle-spining and wavelet packet transform are two special cases.

3.1. Global Thresholding. Noise is assumed to have a Gaussian distribution due to the central limit theorem. Global thresholding assumes that Gaussian noise has the same frequency distribution and amplitude for all orthogonal bases that span the same data space.[17] There are several ways to apply these thresholds to the wavelet coefficients: hard, soft, non-negative garrote, and firm. They are defined as

Hard:

$$x_i^* = \begin{cases} 0 & \text{if } |x_i| \le t \\ x_i & \text{if } |x_i| > t \end{cases} \tag{5}$$

Soft:

$$x_i^* = \begin{cases} 0 & \text{if } |x_i| \le t \\ \text{sign}(x_i)(|x_i| - t) & \text{if } |x_i| > t \end{cases} \tag{6}$$

Garrote:

$$x_i^* = \begin{cases} 0 & \text{if } |x_i| \le t \\ x_i - t^2/x_i & \text{if } |x_i| > t \end{cases} \tag{7}$$

Firm:

$$x_i^* = \begin{cases} 0 & \text{if } |x_i| \le t_1 \\ \text{sign}(x_i) \times t_2(|x_i| - t_1)/(t_2 - t_1) & \text{if } t_1 < |x_i| \le t_2 \\ x_i & \text{if } |x_i| > t_2 \end{cases} \tag{8}$$

for which x_i and x_i* stand for the wavelet coefficients before and after thresholding, respectively.

For the first three methods, the wavelet coefficients are partitioned into two parts by the threshold t. Hard thresholding is a classic way to remove noise and is the only thresholding method whose function is discontinuous (i.e., removes coefficients with low magnitude). Soft thresholding shrinks all large coefficients by the value of the threshold as well as removes all small coefficients.[24] Soft thresholding is analogous to apodization in the Fourier transform methods. Hard thresholding introduces discontinuities into the denoised data but has smaller RMS errors than soft thresholding. Soft thresholding tends to generate denoised data that is continuous at the expense of larger RMS errors. Soft thresholding tends to over-smooth abrupt changes and broaden sharp peaks and may give a visually better estimator. Non-negative garrote thresholding shrinks the large coefficients by a nonlinear continuous function and removes small coefficients.[20,25] Firm thresholding has two thresholds; the

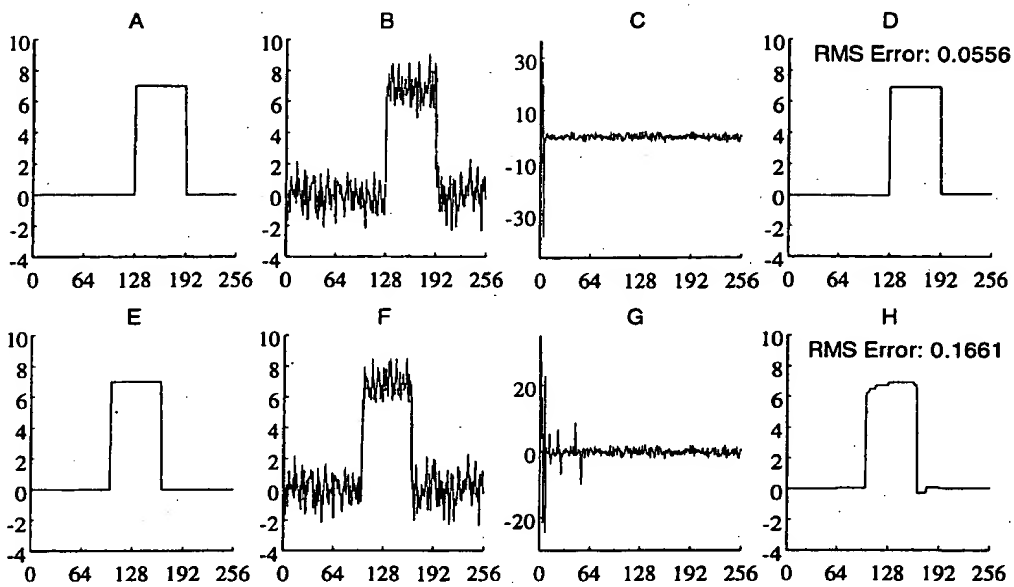

Figure 3. Demonstration of importance of location of discontinuities to wavelet transform. The discontinuities are at points 128 and 192 for the upper signal and points 100 and 164 for the bottom signal. The Haar transform is used.

wavelet coefficients are partitioned into three treatments: (1) retain the large coefficients, (2) remove the small coefficients, and (3) linearly shrink the middle coefficients.[26] Both garrote and firm thresholding methods attempt to moderate the limitations of the hard and soft thresholding methods.

3.2. Level-Dependent Thresholding. Level-dependent thresholding uses different thresholds for each transform level. SURE is usually applied to select thresholds for the coefficients in different levels. Universal soft threshold can also be applied if different levels have different noise values, as calculated by eq 2. SURE does not work well when the wavelet representations are sparse (i.e., contain mostly zero values). SURE has been combined with the universal method to yield a hybrid method that circumvents this problem. The hybrid method uses a sample variance at each level to determine if the representation at that level is sparse. If the level is not sparse, the SURE threshold is used, otherwise a universal threshold is used.

3.3. Data-Dependent Thresholding. Data-dependent threshold (DDT) is determined by a statistical test within each level. The change-point (CP) approach is a data-dependent level-by-level recursive scheme, based on the standard likelihood ratio test. First, all coefficients in a level are assumed to represent noise.Then a test statistic is computed and compared with the critical value. If the test is significant, the largest absolute value is considered non-noise and is removed from the noise coefficients. Using the retaining coefficients, the procedure continually repeats until the test is insignificant.After determining the threshold, which is the maximum of the coefficients tested to be noise, a soft thresholding is performed. Therefore, this method tries to extract a subset of coefficients that behave like pure noise. By adjusting the level α of the hypothesis tests, one can control the smoothness of the resulting estimator. A typical choice (0.01) tends to give a smoother denoised and visually appealing signal but a larger root mean square error (RMS). An unusually large choice (0.999) may give a smaller RMS error.[10] Another CP approach uses the localization property of DWT. Common thresholding methods use the magnitudes of the wavelet coefficients only. Because a sharp peak in the signal results in several nonzero wavelet coefficients that are adjacent to each other, it is possible to use CP approaches to take advantage of this positional information.[11]

3.4. Cycle-Spin Thresholding. DWT is similar to FT denoising in that denoising may introduce artifacts to the regenerated data, especially around some discontinuities such as sharp peaks or abrupt changes in the data. The cycle-spin thresholding denoising method is intended to reduce the artifacts.[27] The data are first cycle-spun (i.e., translated) by h points, transformed and thresholded, transformed back, and spun back by h points to their original position. Spinning refers to translating the data with the points shifted past the zero index added onto the other side of the data object (i.e., rotated). The reason for this transformation is that the artifacts caused by DWT are connected intimately with the actual location of the discontinuity in the data. A demonstration is given in Figure 3. The only difference between the upper signal and the bottom signal is the position of the edges. The spectrum in the panel E is a 28-point-spun version of the spectrum in the panel A. These figures show that the positions of discontinuities are important in DWT denoising methods. Cycle-spinning uses the localization property of the DWT, which the FT does not have. Therefore, by shifting some points in the original spectrum, the wavelet spectrum may change from panel G to panel C. In panel G, some signal coefficients are small and some are even buried into noise coefficients; but in panel C, the signal is represented by a few large coefficients. For a given signal, a best shift h_{opt} may be selected by optimization. A given signal can be realigned to minimize artifacts, but there is no guarantee that this will always be the case. For example, when a signal contains several discontinuities, they may interfere with each other: the best shift for one discontinuity may also be the worst shift for another discontinuity. Therefore, the idea of averaging all shifts, which is called translation invariant (TI) denoising, usually can give a much better result than ordinary cycle-spin denoising. Moreover, there is no guarantee that the TI averaging result is better

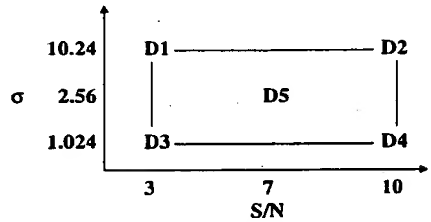

Figure 4. Experimental design of the synthetic data.

than the results of best shift and the uncycled methods. The cycle-spin approach provides a natural way to generate multiple estimators for the same object. However, it is more computationally intensive compared to ordinary DWT denoising.

3.5. Wavelet Packet Transform. Wavelet packet transform (WPT) is another powerful denoising tool.[28,29] WPT is a generalized form of DWT, in which both smooth and detail parts are subject to further transforms. A full transformed matrix containing J (= log N) transform levels is used to search for a best basis. The best basis can be chosen using different criteria. Shannon entropy is a very common one, which is defined as

$$S = -\sum_j p_j \times \log(p_j) \tag{9}$$

for which $p_j = |x_j|^2/||x||^2$, and $p \log p = 0$ for $p = 0$. By comparing the possible combinations of all the wavelet coefficients at the different levels, a best basis can be obtained that is the combination of coefficients x with minimum entropy. The other criteria include (A) minimum $\sum \log|x_j|$, (B) minimum number larger than t, and (C) minimum SURE.

Practically, all of these DWT methods leave intact the last two or three levels, i.e., the four or eight points in wavelet domain spectrum, shown in Figure 2, because they represent the most important information.

EXPERIMENTAL SECTION

Synthetic data were used to evaluate different denoising methods. The data consisted of 256 points that contained a single Gaussian peak. Signal-to-noise ratio (*S/N*) and peak width were the parameters for synthesizing Gaussian peaks. *S/N* specifies the ratio of the peak height to the standard deviation of the Gaussian noise. Five data objects were synthesized using a two level square design with a central composite point, as shown in Figure 4. *S/N* ranged from 3 to 10, and peak widths ranged from 1.024 to 10.24 in units of data point. Peak width refers to the standard deviation for the Gaussian function. The five synthetic data objects are displayed in Figure 5.

An infrared absorbance spectrum of sunflower oil was acquired from a Perkin Elmer Model 1600 FTIR spectrophotometer equipped with a DTGS detector and a KBr beam splitter. The spectrum was signal-averaged 256 times in the range 450-4400 cm^{-1} at 2 cm^{-1} resolution. This spectrum was used as a true signal, and some white Gaussian noise was added. The noise level was 5% of the maximum peak height. The figure of merit was the RMS error between the absorbance spectrum before adding noise to it and the denoised spectrum. The IR spectrum and the signal with noise added are given in Figure 5.

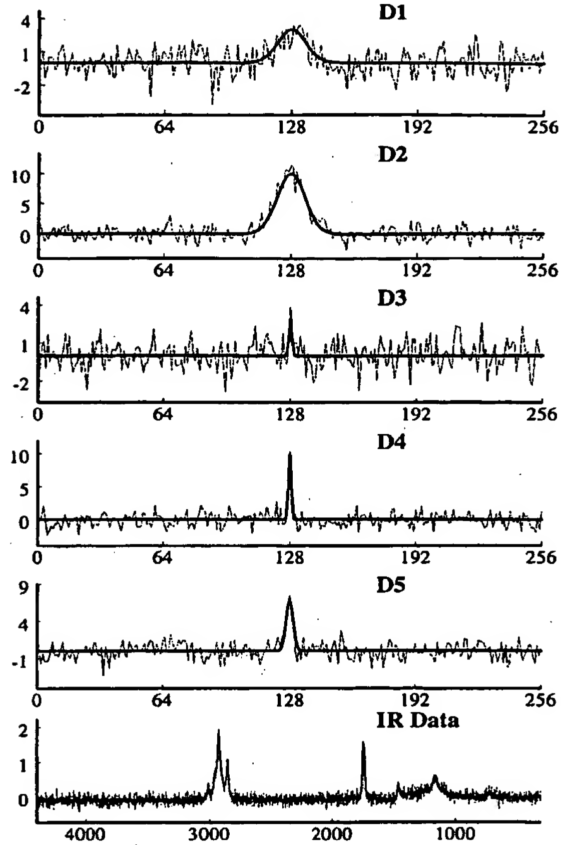

Figure 5. The synthetic data objects.

A real data object was a chromatogram acquired from the injection of a standard solution of morphine and nalorphine. The column was a Supelcosil ABZ+ with inner diameter of 2.1 mm. The mobile phase was 90:9:1 of sodium hexametaphosphate (0.01 M, pH = 3.8): methanol:THF with flow rate of 0.6 mL/min. The concentration of morphine was 1.0 ng/mL and the injection volume was 20.0 μL. The detector was a Shodex CL-2 (JM Science) chemiluminescence detector. Experimental details were described elsewhere.[30] The chromatogram contained 1024 points, and the sampling frequency for the chromatographic data was 10 Hz.

The four threshold selection methods can be combined with the global thresholding methods, level-dependent thresholding methods, data-dependent thresholding method, TI method, and WPT thresholding methods. Theoretically, many choices for a single wavelet transform denoising are available. Actually, most methods yield similar results and some combinations of threshold selection and shrinkage methods do not have a theoretical basis. In this work, the following 12 denoising methods were examined:

1. UNIVERSAL: universal threshold, global hard thresholding
2. MINIMAX-HARD: minimax threshold, global hard thresholding
3. MINIMAX-SOFT: minimax threshold, global soft thresholding
4. MINIMAX-GARROTE: minimax threshold, global non-negative garrote thresholding

5. MINIMAX-FIRM: minimax threshold, global firm thresholding
6. MDL: MDL threshold, global hard thresholding
7. SOFT: universal threshold, global soft thresholding
8. MULTI-SURE: SURE threshold, level-dependent hard thresholding
9. MULTI-HYBRID: SURE + universal threshold, level-dependent hard thresholding
10. DDT: data dependent threshold, level-dependent hard thresholding
11. TI: universal threshold, global hard TI thresholding
12. WPT: universal threshold, global hard WPT thresholding

The root mean square difference between the true signal and the regenerated signal that was obtained from noisy data was used to evaluate these denoising methods. First, 22 wavelet types were assembled for evaluating the denoising methods. The denoising methods were evaluated with the suite of filters (i.e., wavelet types such as symmlet 8) by the RMS error. The denoised data was compared to the true signal.

The data processing computations were performed with Matlab version 5.2 on an Indigo2 Impact 10000 195MHz SGI workstation equipped with 192MB of RAM, which was operated under IRIX 6.2 operation system. All wavelet filter coefficients and some programs were obtained from WaveLab package.[31] The results were transferred to a PC, and the figures were generated with Axum 5.0B for Windows (Mathsoft Inc.).

RESULTS AND DISCUSSION

With the synthetic data the underlying signal is known, so the accuracy of the denoising method may be quantified using RMS error. This error is a measure of the disparity between the denoised signal and the underlying signal. The real data are somewhat subjective and can only be evaluated through visual inspection.

1. The Denoising Ability of DWT Denoising Methods. The results from the synthetic data are given in Table 1. Relative RMS error (RRMS) for each method is reported. RRMS is the ratio of the RMS error and the maximum signal, and it estimates how much noise is suppressed. The values for the data in this table represent minimum errors of the optimized parameters. The minimax results were computed by using the modified threshold.[19] For the DWT-based denoising methods, the minimum RMS error was obtained from 22 filters. The Savitzky–Golay and FT denoising results are given in Table 1. The best results for the Savitzky–Golay is the minimum RMS error obtained from 5 to 71-point cubic filters. For the FT, a trapezoidal apodization function was used. The result was exhaustively computed from a combination of all apodization and truncation frequencies.

From the table, one can see that the DWT-based methods can be classified into three groups according to their performances. The first group gives large RRMS errors, so they are not suitable for analytical data. This group includes SOFT, MINIMAX-SOFT, and MULTI-SURE methods. The SOFT method tends to over-smooth noisy data; it retains the signal feature but introduces artifacts that furnish a larger RRMS error. The MINIMAX-SOFT method is similar, although the results are better than SOFT method. The MULTI-SURE method is too conservative with respect to filtering and retains lower frequency noise.

Table 1. Minimum RRMS of Different Denoising Methods for the Synthetic Data (%)

method	D1	D2	D3	D4	D5	IR
no denoising	34.47	9.92	34.23	9.63	14.36	4.93
FT	9.83	3.26	7.13[a]	6.38	7.90	1.53
Savitzky–Golay	9.53	3.35	8.17[a]	6.56	7.94	1.48
UNIVERSAL	7.67	3.22	8.77	3.19	3.84	1.29
MINIMAX-HARD	7.67	3.24	7.50	3.02	3.84	1.26
MINIMAX-SOFT	9.97	3.47	7.63	3.85	5.34	1.80
MINIMAX-GARROTE	9.00	3.04	7.47	3.41	4.56	1.36
MINIMAX-FIRM	9.03	3.01	7.43	3.48	4.63	1.27
MDL	9.17	3.22	7.13[a]	3.81	4.01	1.63
SOFT	11.93	4.98	7.50[a]	4.77	7.34	2.41
MULTI-SURE	19.00	4.80	16.07	4.47	7.17	1.88
MULTI-HYBRID	7.67	2.64	7.20	4.31	5.69	1.25
DDT	8.90	2.77	9.50	3.66	5.66	1.43
TI	6.77	2.03	7.30	2.57	3.91	1.00
WPT	4.07	2.75	10.00[a]	1.64	2.37	1.35

[a] These methods treat this data as pure noise. The denoising results are just a baseline.

The second group includes the UNIVERSAL, MINIMAX-HARD, MINIMAX-GAROTE, MINIMAX-FIRM, MDL, MULTI-HYBRID, and DDT methods. They have similar denoising abilities and usually can give better results than the FT and SG methods, especially for the data objects D3 and D4 that contain relatively narrow peaks. The remaining TI and WPT belong to the third group. They usually can give the best denoising results. Therefore, the second and third groups of methods are effective for denoising analytical data. All optimal denoising results for data D5 (i.e., central design point) are given in Figure 6.

2. Practical Considerations of the WT Denoising Methods. Note the RRMS for each WT-based method in Table 1 is the best result among the 22 filters, which is possible only if the underlying signal is known. For real data, there is no such criterion to select a filter. These WT denoising methods are applicable to real data. The question remains on which filters will furnish good results. Among these WT-based methods, only the MDL method can select a filter according to the MDL values. All others denoising methods have no ability to select a filter other than some empirical rules that are based on the data. For some types of data no rules may exist. Figure 7 plots RMS error for data D5 with respect to the 22 filters for the four denoising methods: UNIVERSAL, MDL, TI, and WPT. The underlying signal has a peak intensity of 7.

The WPT is very dependent on the filter, very small errors may be obtained for some filters, and the errors may be exceedingly large for other filters. Furthermore, this extreme behavior may occur even if the filter types are similar, e.g., symmlet 9 and symmlet 10, which is also true for all other DWT methods except for TI. Therefore, from the viewpoint of a user, there is no reason to select a specific filter until an applicable criterion is chosen, and it is impractical to apply all filters and visually determine the best result.

The MDL threshold selection method is the only one method that selects both a threshold and a filter type. The MDL method selection of wavelet filter was evaluated by the RMS error between the denoised and true signal for 100 synthetic data objects. The true signal was the IR spectrum

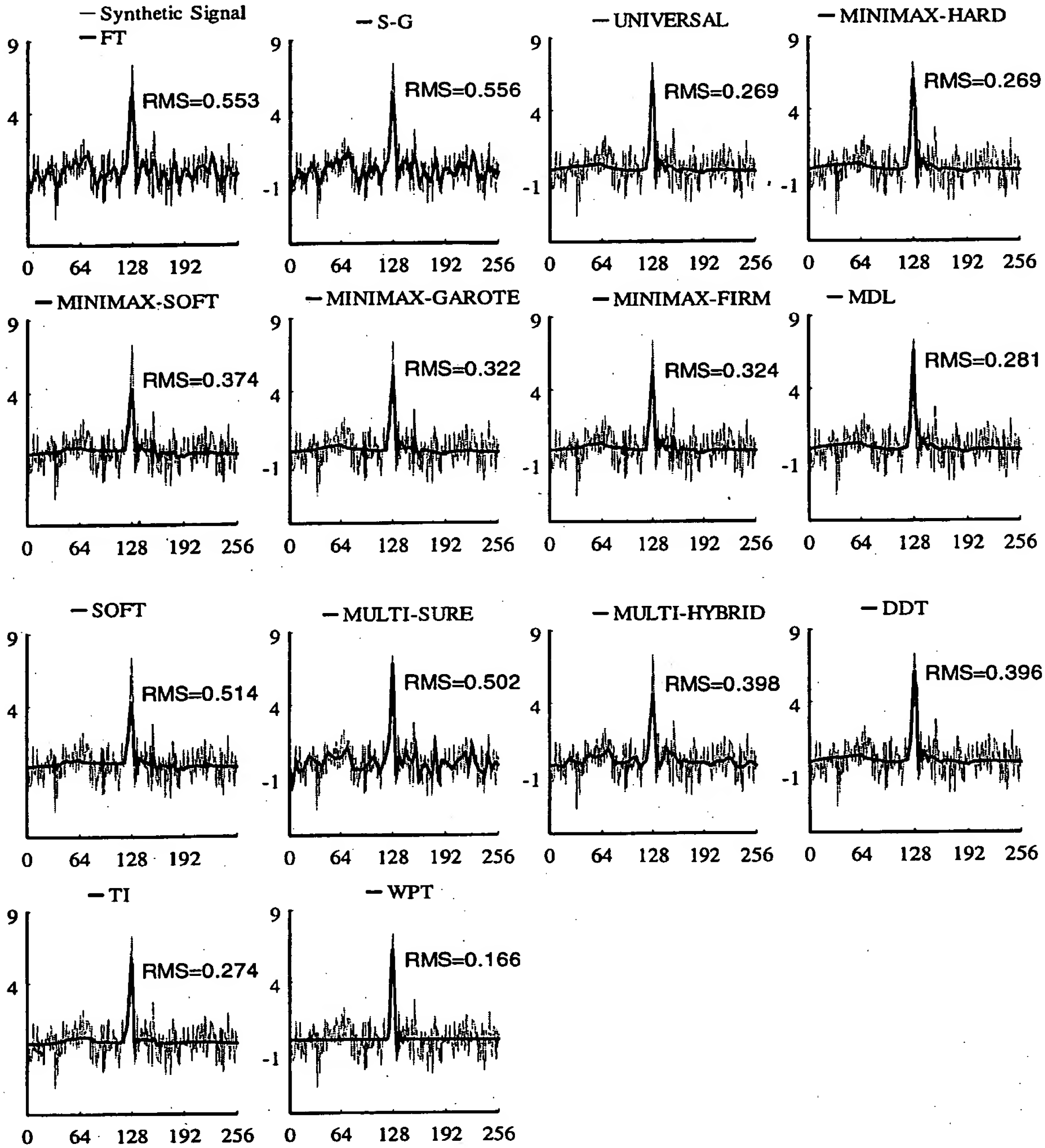

Figure 6. Denoising results for data, the synthetic noisy spectra are shown as well. The result with minimum RMS error is plotted for each method.

in Figure 5, and 100 different white, Gaussian noise spectra were added to the spectrum at a *S/N* of 20. For each noisy IR spectrum, all 22 filters were performed, and the RMS errors were calculated, then it was easy to determine if the MDL selected the best filter. It was found that in 42% cases the MDL chose the filter that furnished the minimum RMS error. In 20% of the cases, the MDL chose the filter that yielded the second smallest RMS error, and in 18% of the cases, the MDL chose the filter that yielded the third smallest RMS error. In 91% cases, the MDL selected the filter whose RMS error was in the five lowest errors among all 22 filters. The difference between the fifth smallest RMS error and the smallest error did not differ significantly. One example is given in Figure 8. In the figure, RMS errors were arranged in ascending order, so it is easy to compare the difference of RMS errors. The spectrum had a maximum of 2 absorbance units. The worst case (spectrum 74) is also given in Figure 8, and the RMS error is about the average. These results indicate MDL is a very good method for selecting a wavelet filter, as opposed to arbitrarily selecting one. In Figure 7, the diamonds indicate which filter MDL selected for the synthetic data.

From Figure 7, for the TI method, the RMS curves are relatively flat, especially for the coiflet and symmlet transform families. This method is filter-insensitive, because it gives equivalent denoising regardless of the filter that is used. These denoising results are typically better than most other DWT methods. For example of denoising the noisy IR data, the TI gives the minimum RMS error of 0.0191 when using symmlet 5 and gives the maximum RMS error of 0.0220

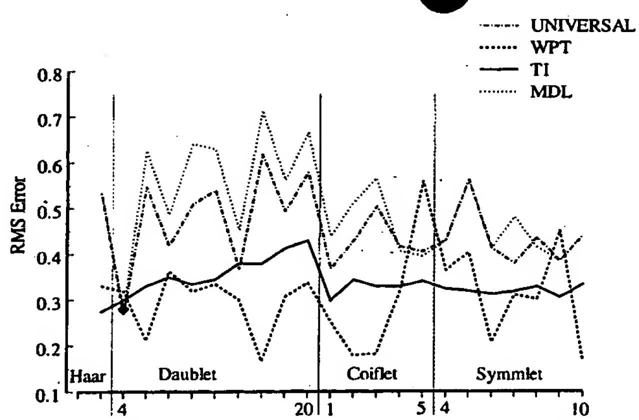

Figure 7. RMS error of denoising with respect to the filter type.

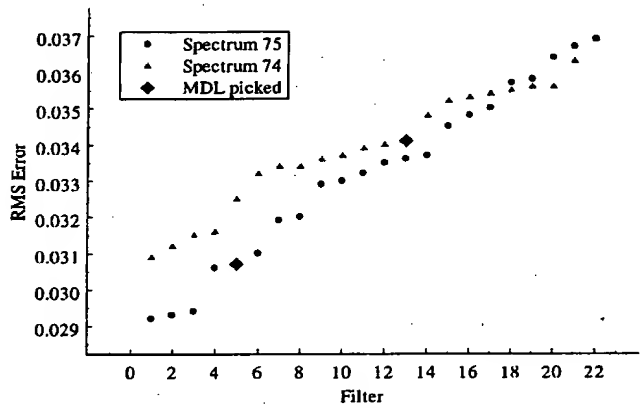

Figure 8. An example in which the MDL chose the worst filter for the spectrum 74. Another example is the MDL chose the fifth minimum RMS. The MDL chooses the fifth or better filter with 91% possibility. The plot is arranged in ascendant RMS order. The x-axis indicates different filters.

when using daublet 20. Even the maximum RMS error is lower than the RMS error from other methods.

The MDL and TI are two practical denoising methods and generate the good results from the synthetic data evaluations. HPLC data was used to test these two denoising methods.The chromatogram is given in Figure 9. These data are similar to the synthetic data 1, the peak is not sharp, and it should be easy to denoise for every denoising method. The purpose for choosing this data is to demonstrate the DWT denoising methods at least are comparable to the traditional denoising for a typical analytical signal. The small and large peaks are from the analytes morphine and nalorphine, respectively. The peak of interest is at a retention time of 75 s and is very noisy due to the low concentration of the analyte. The TI and WPT denoising results are given in Figure 10. For comparison, Savitzky–Golay and FT results are also given. For Savitzky–Golay method, an unusual 45-point cubic filter is used. For the FT method, an optimal filter is used. First, the first 200 points, middle 100 points, and last 212 points in the data are considered as noise, the noise power spectrum is calculated and then normalized by the data number. The signal power spectrum of entire data is also computed and normalized. The first point at which the power of signal is lower than the corresponding power of noise is assigned as the cutoff point. Fourier coefficients with higher frequency than the cutoff point are zero-filled. Lastly the inverse FT is applied.

The MDL method automatically picks a suitable filter. For TI, the noise was estimated by eq 2, and symmlet 8 transform was used. The global hard thresholding with universal threshold was used. By visual comparison, the MDL and TI methods are comparable to FT and Savitzky–Golay methods. The advantage for the MDL and TI methods includes the nonarbitrary criteria, which means their denoising criteria apply to almost all data. Another advantage for DWT denoising includes fast computation, the computation for a complete DWT is comparable to FFT, which is important if a large amount of data have to be denoised on a slow computer.

CONCLUSIONS

DWT provides a variety of denoising methods, which is advantageous and disadvantageous. A certain method may be suitable for specific types of data, although two similar wavelets may yield significantly different denoising results. Problems may arise if improper wavelet filters are chosen. After evaluating the different methods by using synthetic

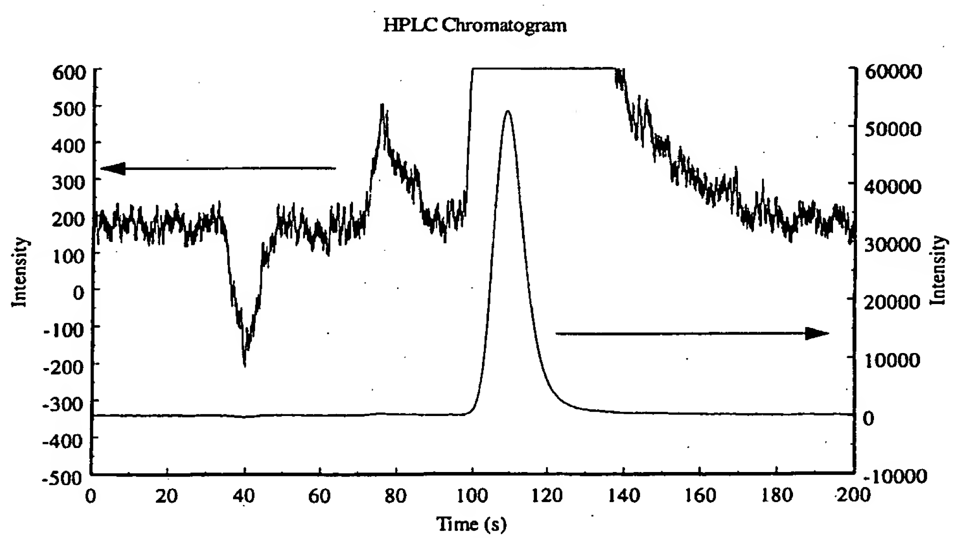

Figure 9. Real HPLC data. It is very noisy for the smaller morphine peak in front of the larger metabolite peak.

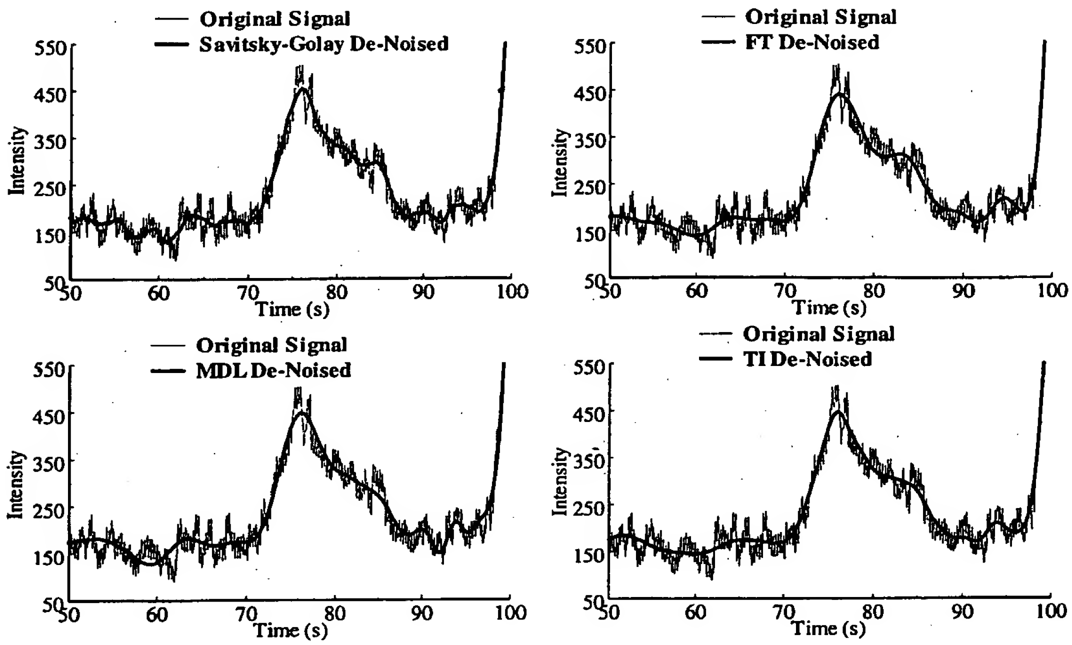

Figure 10. Denoising results for the HPLC data.

data, we advocate the use of translation invariant (TI) and minimum description length (MDL) denoising methods, which are practical and generate better results. These two methods are more objective compared to most denoising methods. The MDL method automatically selects a filter type and a threshold. Furthermore, MDL method does not require a priori knowledge of the noise level, which for some cases may be difficult or impossible to estimate. The TI method reduces the artifacts, and the results are almost filter-independent. However, the TI is the more computationally intensive method.

Compression of analytical data is re-emerging as an important research area, as chemical sensors are becoming smaller, less expensive, and more prevalent. The sensor data may have to be stored on miniaturized devices, which may limit storage capacity, or transmitted which may have limited bandwidth. Wavelet transforms offer a potential means for compressing measurement data and removing unwanted noise. In some cases, wavelets may offer advantages over the standard method of Fourier compression. To achieve these advantages, it is essential that the correct wavelet filter be selected.

ACKNOWLEDGMENT

Dr. Anthony Andrews is thanked for providing the HPLC data. Chuanhao Wan, Tricia Buxton, Susan Slagel, and Ronald Tucceri are thanked for their helpful suggestions. This work was presented in part at the 24th Annual Conference of the Federation of Analytical Chemistry and Spectroscopy Societies, Providence, RI.

REFERENCES AND NOTES

(1) Bruce, A.; Donoho, D.; Gao, H. Wavelet Analysis. *IEEE Spectrum* **1996**, *33(10)*, 26–35.

(2) Jouan-Rimbaud, D.; Walczak, B.; Poppi, R. J.; de Noord, O. E.; Massart, D. L. Application of Wavelet Transform to Extract the Relevant Component from Spectral Data for Multivariate Calibration. *Anal. Chem.* **1997**, *69*, 4317–4323.

(3) Barclay, V. J.; Bonner, R. F.; Hamilton, I. P. Application of Wavelet Transforms to Experimental Spectra: Smoothing, Denoising, and Data Set Compression. *Anal. Chem.* **1997**, *69*, 78–90.

(4) Walczak, B.; Massart, D. L. Wavelet Packet Transform Applied to a Set of Signals: A New Approach to the Best-Basis Selection. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 39–50.

(5) Walczak, B.; Massart, D. L. Noise Suppression and Signal Compression Using the Wavelet Packet Transform. *Chemom. Intell. Lab. Syst.* **1997**, *36*, 81–94.

(6) Alsberg, B. J.; Woodward, A. M.; Kell, D. B. An Introduction to Wavelet Transforms for Chemometricians: A Time-Frequency Approach. *Chemom. Intell. Lab. Syst.* **1997**, *37*, 215–239.

(7) Daubechies, I. *Ten Lectures on Wavelets*; SIAM: Philadelphia, PA, 1992.

(8) Donoho, D. L.; Johnstone, I. M.; Kerkyacharian, G.; Picard, D. *J. Roy. Stat. Soc. Ser. B* **1995**, *57*, 301–369.

(9) Donoho, D. L.; Johnstone, I. M. Adapting to Unknown Smoothness via Wavelet Shrinkage. *J. Am. Statist. Assoc.* **1995**, *90(432)*, 1200–1224.

(10) Ogden, T.; Parzen, E. Data Dependent Wavelet Thresholding in Nonparametric Regression with Change-point Applications. *Computational Statistics Data Analysis* **1996**, *22*, 53–70.

(11) Ogden, T.; Parzen, E. Change-point Approach to Data Analytic Wavelet Thresholding. *Statistics Computing* **1996**, *6*, 93–99.

(12) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.

(13) Steiner, J.; Termonia, Y.; Deltour, J. Comments on Smoothing and Differentiation of Data by Simplified Least Squares Procedure. *Anal. Chem.* **1972**, *44*, 1906–1909.

(14) Madden, H. H. Comments on the Savitzky–Golay Convolution Method for Least Squares Fit Smoothing and Differentiation of Digital Data. *Anal. Chem.* **1978**, *50*, 1383–1386.

(15) Bush, C. A. Fourier Method for Digital Data Smoothing in Circular Dichroism Spectrometry. *Anal. Chem.* **1974**, *46*, 890–895.

(16) Maldacker, T. A.; Davis, J. E.; Rogers, L. B. Applications of Fourier Transform Techniques to Steric-Exclusion Chromatography. *Anal. Chem.* **1974**, *46*, 637–642.

(17) Donoho, D. L. In *Proceedings of Symposia in Applied Mathematics: Different Perspectives on Wavelets*; Daubechies, I., Ed.; American Mathematical Society: Providence, RI, 1993; Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data pp 173–205.

(18) Donoho, D. L.; Johnstone, I. M. Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika* **1994**, *81*, 425–455.

(19) Bruce, A. G.; Gao, H. Understanding Waveshrink: Variance and Bias Estimation. *Biometrika* **1996**, *83(4)*, 727–745.

(20) Gao, H. Preprint: Wavelet Shrinkage Denoising Using the Non-Negative Garrote; SoftMath, Inc.: 1997.

(21) Stein, C. Estimation of the Mean of A Multivariate Normal Distribution. *Annals Statistics* **1981**, *9(6)*, 1135−1151.

(22) Gao, H. Preprint: Threshold Selection in WaveShrink; SoftMath, Inc.; 1997.

(23) Saito, N. In *Wavelets in Geophysics*; Foufoula-Georgiou, E., Kumar, P., Eds.; Academic Press: San Diego, CA, 1994; Simultaneous Noise Suppression and Signal Compression Using A Library of Orthonormal Bases and the Minimum Description Length Criterion, pp 299−324.

(24) Donoho, D. L. Denoising by Soft-Thresholding. *IEEE Trans. Inf. Theory* **1995**, *41(3)*, 613−627.

(25) Breiman, L. Better Subset Regression Using the Nonnegative Garrote. *Technometrics* **1995**, *37*, 373−384.

(26) Gao, H.; Bruce, A. G. Waveshrink with Firm Shrinkage. *Statistica Sinica* **1997**, *7(4)*, 855−874.

(27) Coifman, R. R.; Donoho, D. L. In *Wavelets and Statistics*; Antoniadis, A., Oppenhein, G., Eds.; Springer-Verlag: New York, 1995; Translation-Invariant Denoising, pp 125−150.

(28) Coifman, R. R.; Wickerhauser, M. V. Entropy-based Algorithms for Best Basis Selection. *IEEE Trans. Inf. Theory* **1992**, *38(2)*, 713−718.

(29) Berger, J.; Coifman, R. R.; Goldberg, M. J. Removing Noise from Music Using Local Trigonometric Bases and Wavelet Packets. *J. Audio Eng. Soc.* **1994**, *42*, 808−818.

(30) Amiott, E.; Andrews, A. R. J. Morphine Determination by HPLC with Improved Chemiluminescence Detection Using A Conventional Silica Based Column. *J. Liq. Chrom. Rel. Technol.* **1997**, *20*, 311−325.

(31) Buckheit, J. B.; Donoho, D. L. In *Wavelets and Statistics*; Antoniadis, A., Oppenhein, G., Eds.; Springer-Verlag: New York, 1995. Wavelab and Reproducible research, pp 55−81. WaveLab 7.01 is available at http://playfair.stanford.edu/~wavelab.

CI980210J